DOCUMENT RESUME

ED 241 495 SP 023 948

AUTHOR Yinger, Robert J.; Clark, Christopher M.
TITLE Self-Reports of Teacher Judgment. Research Series No. 134.
INSTITUTION Michigan State Univ., East Lansing. Inst. for Research on Teaching.
SPONS AGENCY National Inst. of Education (ED), Washington, DC.
PUB DATE Jul 83
CONTRACT 400-81-0014
NOTE 30p.; Paper presented at the Annual Meeting of the American Educational Research Association (Montreal, Canada, April 11-15, 1983).
AVAILABLE FROM Institute for Research on Teaching, College of Education, Michigan State University, 252 Erickson Hall, East Lansing, MI 48824 ($3.00).
PUB TYPE Speeches/Conference Papers (150) -- Reports - Research/Technical (143) -- Information Analyses (070)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Classroom Research; *Cognitive Processes; *Decision Making; Elementary Education; *Elementary School Teachers; *Evaluative Thinking; *Instructional Materials; Language Arts; Media Selection; Metacognition; Research Methodology; *Teacher Attitudes; Teacher Characteristics; Teaching Experience

ABSTRACT

Comparison was made of the results of three analyses of teacher judgments concerning the selection of curriculum materials for the teaching of writing in elementary school. Twenty-five male and female fourth and fifth grade teachers, with teaching experience ranging from 4 to 33 years, responded to questions on their judgments of the value of language arts activities (described by short statements of purpose and a listing of the steps involved in planning and conducting the activity). The intent of the study was to determine the validity of three types of research used in exploring the judgment process: (1) policy capturing analysis; (2) process tracing analysis; and (3) analysis of teachers' self-reports of their judgment processes. Among conclusions reached were the following: (1) Teachers as judges may have better insight into their own decision processes than researchers usually give them credit for. Closer attention should be paid to differences in language and level of detail offered by the various methods and to what kind of data is used to evaluate the validity of verbal reports. (2) Better models of the tasks in which judgment is being examined should be developed. (3) More should be known about how experience influences judgment. (4) Multi-method approaches will probably provide more accurate results. (JD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED241495

Research Series No. 134

SELF-REPORTS OF TEACHER JUDGMENT

Robert J. Yinger and Christopher M. Clark

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
Christopher M. Clark
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☐ This document has been reproduced as received from the person or organization originating it.
☒ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

Institute for Research on Teaching

College of Education — Michigan State University

CS 203 948

Research Series No. 134

SELF-REPORTS OF TEACHER JUDGMENT

Robert J. Yinger and Christopher M. Clark

Published By

The Institute for Research on Teaching
252 Erickson Hall
Michigan State University
East Lansing, Michigan 48824

July 1983

This work is sponsored in part by the Institute for Research on Teaching, College of Education, Michigan State University. The Institute for Research on Teaching is funded primarily by the Program for Teaching and Instruction of the National Institute of Education, United States Department of Health, Education, and Welfare. The opinions expressed in this publication do not necessarily reflect the position, policy, or endorsement of the National Institute of Education. (Contract No. 400-81-0014)

Institute for Research on Teaching

The **Institute for Research on Teaching** was founded at Michigan State University in 1976 by the National Institute of Education. Following a nationwide competition in 1981, the NIE awarded a second contract to the IRT, extending work through 1984. Funding is also received from other agencies and foundations for individual research projects.

The IRT conducts major research projects aimed at improving classroom teaching, including studies of classroom management strategies, student socialization, the diagnosis and remediation of reading difficulties, and teacher education. IRT researchers are also examining the teaching of specific school subjects such as reading, writing, general mathematics, and science, and are seeking to understand how factors outside the classroom affect teacher decision making.

Researchers from such diverse disciplines as educational psychology, anthropology, sociology, and philosophy cooperate in conducting IRT research. They join forces with public school teachers, who work at the IRT as half-time collaborators in research, helping to design and plan studies, collect data, analyze and interpret results, and disseminate findings.

The IRT publishes research reports, occasional papers, conference proceedings, a newsletter for practitioners, and lists and catalogs of IRT publications. For more information, to receive a list or catalog, and/or to be placed on the IRT mailing list to receive the newsletter, please write to the IRT Editor, Institute for Research on Teaching, 252 Erickson Hall, Michigan State University, East Lansing, Michigan 48824-1034.

Co-Directors: Jere E. Brophy and Andrew C. Porter

Associate Directors: Judith E. Lanier and Richard S. Prawat

Editorial Staff
Editor: Janet Eaton
Assistant Editor: Patricia Nischan

Abstract

This report compares the results of three analyses of data on teachers' judgments concerning the selection of curriculum materials for the teaching of writing in elementary school. The three analyses are a policy capturing analysis, a process tracing analysis, and an analysis of teachers' self-reports of their judgment processes. The authors conclude that teachers as judges may have better insight into their own decision processes than researchers have given them credit for. They call for more attention in judgment research to modeling of the judgment task and hypothesizing about the differences in judgment heuristics used by novice and experienced teachers.

SELF-REPORTS OF TEACHER JUDGMENT[1]

Robert J. Yinger and Christopher M. Clark[2]

Recently, it has been reasserted that teachers, like people in general, are unaware of how they use and weigh information to make judgments (Shavelson and Stern, 1982). This claim is based on a comparison of teachers' self-reports of their judgment "policies" with mathematical models of these decisions generated in policy-capturing research.

Within the research on human judgment there has been debate over whether these findings are due to a person's inability to properly perceive mental processes (e.g., Nisbett and Wilson, 1977) or are due to the inherent characteristics of the linear models used in policy capturing (e.g., Dawes and Corrigan, 1974). To date, evidence has been most frequently collected by comparing policy capturing models and self-reports on the basis of their predictive ability. This paper reports an effort to extend researchers knowledge about a judge's self- (or meta-cognitive) awareness of mental processes by comparing self-reports of teachers' judgments about instructional materials with both a policy-capturing model and a more descriptive process-tracing model.

Policy capturing is a popular and frequently used method of studying and representing human judgment. This approach begins with a simple model

---

[1]This paper was presented at the April, 1983 annual meeting of the American Educational Research Association in Montreal.

[2]Robert Yinger is a former IRT senior researcher with the Teacher Planning Project and an associate professor of education at the University of Cincinnati. Chris Clark coordinated the IRT's Teacher Planning Project and now co-coordinates the Written Literacy Project. He is an associate professor in the Department of Counseling, Educational Psychology, and Special Education.

(usually linear) and attempts to reproduce the inferential response of a particular judge. Of central interest in this paradigm is how judges weigh and combine information provided in the form of discernable cues or features of the objects to be judged. This approach has been used recently to study a number of aspects of teacher thinking including teacher judgments about characteristics of effective teachers (Anderson, 1977), classroom organization (Borko, 1978), decisions about reading instruction (Borko & Niles, 1982), classroom management (Cone, 1978), instructional strategies (Russo, 1978), and instructional content (Floden, Porter, Schmidt, Freeman, & Schwille, 1981).

Process tracing methods of studying judgment take a very different approach to the problem of investigating and representing thinking processes. Since introspective reports of many judges seem to indicate the presence of a complex, configural judgment process, process-tracing methods begin with a complex representation of the judgment in the form of verbal protocols and attempt to simplify the processes by representing the judgment in the form of decision trees, network representations, or flow diagrams. This approach has been used most widely in cognitive psychology (especially the study of problem solving) and has been only rarely applied to the study of teacher thinking.

This study is one part of a series investigating teacher judgment during the evaluation of instructional materials. This series includes a study investigating the factors influencing the selection of instructional activities (Clark, Yinger, & Wildfong, 1978), a policy-capturing study of teacher judgment (Yinger, Clark, & Mondol, 1981), a process tracing study of teacher judgment (Yinger & Clark, 1982), and an analysis of teachers' self-reported judgment processes (this paper).

The underlying hypothesis of these studies is that the selection of attractive, appropriate, and effective instructional activities is an

important step in teacher planning for instruction (Yinger, 1977). We have argued elsewhere (Clark & Yinger, 1977) that a greater number and variety of studies are required about teacher judgment of students, of curriculum materials, and of other important aspects of the classroom environment before such research will be useful in policy and training decisions. This set of studies adds to the teacher judgment data by investigating teacher thinking in realistically complex situations. By applying various modeling methods to judgment situations like those regularly encountered by elementary school teachers, we are also evaluating the usefulness of these methods for describing the complexities and subtleties of teachers' mental lives.

Method

Subjects

The subjects in this study were 25 fourth and fifth grade teachers from two Michigan school districts. Eight of the teachers were male and 17 were female. Their ages ranged from the mid-20s to the mid-30s. The average number of years of teaching experience was 9 years with a range from 4 to 33 years. Sixteen teachers taught in self-contained classrooms, while nine taught in team-teaching situations or a combination of team-teaching and departmental arrangements. Seven of the teachers taught in urban settings, 18 in suburban communities, and one in a rural area. All of the teachers volunteered for the study and were paid for their participation.

Materials

The materials for the study consisted of 32 one- or two-page descriptions of language-arts writing activities. These descriptions were derived from activities selected from a commercially available instructional catalogue of language-arts activities for upper elementary classrooms (Forte, Frank, &

McKenzie, 1973). The activity descriptions were all presented in the same general format consisting of an activity title, a one- or two-sentence statement of the purpose of the activity, and a listing of the steps involved in planning and conducting the activity.

Each activity description was edited to reflect five dimensions found to be important in teachers' judgments of the quality of language-arts instructional materials (Clark, Yinger, & Wildfong, 1978). These dimensions, or cues, were

1. student involvement,
2. difficulty for students,
3. integration with other skills or subject matter,
4. demand on teachers, and
5. fit between stated purpose and described instructional process.

The 32 activity descriptions were constructed to represent a full factorial matrix of high and low values for each cue. The manipulation and final assessment of each description was accomplished by means of independent ratings of each activity by four researchers, with negotiation between raters when disagreement occurred.

We asked each participant to respond to four questions about each activity. On the back of each activity description the questions were stated along with a nine-point continuum to record each response. The four questions or judgments to be made about each activity were as follows:

1. How attractive is this activity to you?
2. How appropriate is this activity as part of a catalogue of language-arts activities for fourth and fifth grade teachers?
3. How likely would you be to use this activity as it is in your present classroom?
4. How effective do you think this activity would be for your students?

Procedure

Having received an explanation of the purposes, procedures, and materials for the study and having completed a set of six warm-up activities, each teacher responded to the four questions for each of the 32 activity descriptions. Nineteen of the teachers (the Policy-Capturing-Only Group) provided data only in the form of their ratings on the reverse side of each activity description. The remaining six teachers (the Process-Tracing Group) participated in individual sessions. These teachers, in addition to recording their ratings for each description, were asked to "think aloud" as they participated in the task. These verbalizations were tape recorded and later typed into protocols of the judgment task. At the conclusion of the judgment sessions, all teachers were asked to respond to an instrument requesting them to report the factors influencing their judgments by distributing 100 points among the general categories: students, self as teacher, materials, and other (could be specified by the teacher).

Data Analysis

Linear regression equations were computed for each of the four judgments made by each participant. The five activity features were treated as independent variables onto which the ratings given to each case were regressed. The regression equations produced by this analysis yielded for each teacher a set of weightings representing his or her cue use for each of the four judgments. For the teachers in the Process-Tracing Group, models of the judgment process were constructed from the verbal protocols.

Results

This study was designed to provide data comparing teachers' self-reports of factors (cues) they considered when judging instructional activities to

cues suggested by policy-capturing and process-tracing representations of judgment. The typical approach to examining self-reports has been to compare the "subjective" weights that they generate to the weights generated by the mathematical models. We will use the cue usage and process information from the process tracing to help interpret the similarities and discrepancies between the other two data sources.

Self-Reports

Self-reports were solicited from all 25 teachers participating in the study. We have separated the reports from the Policy-Capturing-Only Group (N = 19) and the Process-Tracing Group (N = 6) for comparability, since we have all three sources of data for only a portion of the total group.

As mentioned earlier, each teacher was asked at the end of the judgment session to distribuce a total of 100 points across the four categories of "students," "self," "materials," and "other" (which could be specified) for each of the four judgment questions. The distributions reported by each teacher are shown in Table 1.

Table 1 indicates that Students and Self were by far the most heavily weighted factors, garnering 82% of the grand total for the Policy-Capturing-Only Group and 75% for the Process-Tracing Group. The only distinct difference between the self-reports for the two groups was the tendency in the Process-Tracing Group to assign more weight to the "other" category. This difference is primarily due to relatively heavy weighting by two of the teachers in the Process-Tracing Group (T21, T22) on the second judgment (appropriateness). The factors entered by these two teachers were "the discipline of language arts" and "skills taught (by the activity)."

Table 1

Self-Reported Weightings for 25 Teachers Across 4 Judgments

	Judgment																			
	J1				J2				J3				J4				Grand Total			
Teacher	Stu	Slf	Mat	Oth*	Stu	Slf	Mat	Oth	Stu	Slf	Mat	Oth	Stu	Slf	Mat	Oth	Stu	Slf	Mat	Oth
									Policy-Capturing Group (N = 19)											
1	70	20	10		100				100				100							
2	50	30	20		80	10	10		50	25	25		50	50						
3	20	35	20	25	30	10	20	40	20	30	20	30	25	10	25	40				
4	50	25	25		50	10	40		75	10	15		75	10	15					
5	90	10			90	10			90	10			100							
6	40	20	40		10	25	60	5	50	25	25		90	5	5					
7	50	30	20		75	10	15		25	50	25		35	35	30					
8	50	30	20		20	60	20		60	20	20		50	20	30					
9	20	60	20		20	20	30		80	10	10		70	20	10					
10	50	25	25		50	25	25		25	50	25		25	50	25					
11	30	50	10	10	50	40	10		45	30	25		40	30	15	15				
12		80	20		80		20		90		10		90		10					
13	30	20	30	20	40	20	20	20	40	10	20	30	80	5	5	10				
14	40	50	10		50	40	10		70	10	10	10	35	35	30					
15	25	75			75	25			60	15	15		80	10	10					
16	30	55	15		50	50			60	30	10		90	10						
17	25	50	25		75	20	5		90	10			90	10						
18	30	70			75	10			30	60	10		85	10	5					
19	35	65			40	40	20		30	60	10		70	25	5					
$\bar{X}$	39	42	16	3	58	23	16	3	58	24	14	4	67	18	12	3	56	26	15	3
									Process-Tracing Group (N = 6)											
20	25	75			95	5			75	15	10		100							
21	20	70	10		20			80	30	25	5	40	10	5	75	10				
22	20	60	20		30			70	30	60	10		90	10						
23	60	20	20		50	40	10		40	40	20		100							
24	50	50			75			25	50	25	25		75	25						
25	30	30	40		40	20	20	20	25	25	25	25	30	30	20	30				
$\bar{X}$	34	51	15	0	52	11	5	32	42	31	16	11	67	11	16	6	49	26	13	12

Note: Stu: students, Slf: self, Mat: materials, Oth: others

Policy-Capturing Models

The results of the policy-capturing analysis of the 19 teachers in the Policy-Capturing-Only Group is reported in detail in Yinger, Clark, and Mondol (1981), so we will provide only an overall summary here. In general, the regression equations produced by the policy-capturing analysis did not prove to be good models of the judgment process. Forty-four percent of the equations had no significant weights, and overall, the five activity features that we manipulated accounted for less than one fourth of the variance in the teachers' judgments. The significant models were highly idiosyncratic with no discernable trends in cue use.

The results from the six process-tracing teachers was very similar. Of the 24 policy equations generated, only 9 were statistically significant ($p < .05$). Eight of the significant equations were from two teachers, and the average amount of total variance accounted for by the models (R^2) was only .27 (range = .21 - .37). In other words, the models had virtually no descriptive power for four of the six judges, and when statistically significant, could account for only small portions of the teachers' judgment behavior.

Like the teachers in the Policy-Capturing-Only Group, the process-tracing teachers produced differing policies. The significant equations for this group are illustrated in Table 2. Three cues--Fit, Demand, and Involvement--appear most frequently. (Difficulty appears only twice.) In the table, the components of the models are ordered from most to least heavily weighted.

Processing-Tracing Results

We have not yet analyzed the process-tracing protocols for all six of the teachers in the Process-Tracing Group. The analysis of two teachers is reported in Yinger and Clark (1982). This analysis produced information about

Table 2

Components of the Significant Regression Equations ($p<.05$) for Teachers in the Process-Tracing Group

Teacher	Judgment	Components	R^2
T21	J4	Inv	.21
T22	J1	Fit + Dem + Diff	.37
	J2	Fit	.25
	J3	Fit + Dem	.21
	J4	Fit + Dem	.28
T23	J1	Inv	.23
	J2	Inv + Fit	.26
	J3	Inv + Dem + Fit	.27
	J4	Inv + Dem + Fit + Diff	.36

cue use, processing strategies, and probably most important for this analysis, a description of the way in which these two teachers transformed the judgment task.

Basically, what the process-tracing analysis suggests is that when confronted with an activity-judgment task of this type, these teachers did not judge the activity-as-given, but rather transformed the activity into a form that might actually work in his or her classroom--the activity-to-be-used. This transformation may largely be due to the fact that the activity descriptions constitute plans, and plans imply potential use in a specific context. In fact, much of the mental transformation activities were "contextualization" operations, where the teacher was drawing on and incorporating context-specific knowlege about students, environments, and self. What is eventually judged, then, is not the activity presented on the page, but a modified activity in the mind of the judge.

Applying the Three Sources of Data: A Case Study

Teacher 22 is a 4th grade teacher in a rural school with four years of teaching experience. Her self-reports, like those of the rest of the teachers, emphasize the Students and Self Categories. There is some variation across judgments: A majority of points are assigned to Self in Judgments 1 and 3 (Attractiveness, Likelihood of use); her entry in the "Other" category, "Skills taught (by materials)," captures most of the points in Judgment 2 (Appropriateness); and nearly all of the points were assigned to Students in Judgment 4 (Effectiveness). See Table 3 for a summary.

An examination of the policy-capturing models of Teacher 22 indicates a somewhat different emphasis. As summarized in Table 3, the four significant regression equations assign heaviest weight to the factor Fit (between stated purpose and described instructional process), which was represented in every model and the only factor represented in the model for Judgment 2. Demand (on the teacher) was represented in three of the four models, and Difficulty (for students) was incorporated once.

By comparing the policy-capturing models to the self-reports from this teacher, one might conclude that she tends to overweight the attention she actually pays to student factors and underestimates the attention she places on materials. The R^2s for the regression models account for, on the average, only about 25% of the total variance, so there are likely to be many other factors contributing to the judgments.

The process-tracing analysis may contribute to our understanding of the factors considered by Teacher 22, since the protocols suggest the use of certain cues. During the judgment task this teacher mentioned 22 different cues. Four of the five cues manipulated in the activity descriptions were among those mentioned; Integration was omitted. When judging a single activity,

Table 3

Teacher 22: Three Sources of Data

Self-Reports

	J1	J2	J3	J4
Students	20	30	30	90
Self	60		60	10
Materials	20		10	
Other		70*		

* "skills taught (by materials)"

Policy Models

J1	Fit + Dem + Diff
J2	Fit
J3	Fit + Dem
J4	Fit + Dem

Frequency of Cue Use from Process Tracing

Rank Order:

1. Prerequisite instruction needed
2. Students' task-related ability
3. Fit of stated purpose with activity description
4. Age-level appropriateness
5. Fit with own goals
6. Student interest
7. Student enjoyment

this teacher used a moderate number of cues (mean = 6.81). Seven of the cues accounted for approximately 50% of cue use. These are listed in Table 3.

Process-tracing analysis does not produce a metric comparable to the cue weightings derived from a regression model. We can, however, get a rough indication of the importance of various considerations from the number of times cues were used. (It is important to remember that frequency of use provides no clue as to the importance of a cue in a specific deliberation or how it was used in relation to the other cues considered at the same time.)

An examination of Teacher 22's cue use from the process-tracing analysis seems to complement and explain the data obtained from the policy models and the self-reports. The seven most frequently used cues (in rank order) are 1) prerequisite instruction needed, 2) students' task-related ability, 3) fit of stated purpose with activity description, 4) age-level appropriateness, 5) fit with teacher's own goals, 6) student interest, and 7) student enjoyment. Mindful of the caution stated in the previous paragraph, this information may help resolve some of the discrepancy between the policy models and the self-reports.

The cue-use data obtained from the process analysis supports the cue use suggested by the other two data sources. The fact that six of the seven process-tracing cues involve deliberations about students and self (cues ranked 1,2,4,5,6,7) supports the teacher's impression of heavy weighting of the Student and Self Categories of the self-reports. All seven of the cues imply a consideration of the materials at hand. The suggestion that the judgment process is preceded by extensive transformational and contextualization activities also supports an emphasis on students and self--the participants in the preparation and implementation of the activities.

The information from the process analysis also supports the data from the policy-capturing models. Fit, the cue weighted most heavily, was the third most frequently used cue in the process descriptions. Demand on teacher, used in three of the regression models, was also emphasized in the process analysis, assuming that the need for prerequisite instruction implies a demand on the teacher to supply it.

To summarize, the three sources of data seem to provide complementary, though somewhat different, information about this teacher's judgment. The process model seems to be the richest source of data, because it is not restricted to five cues determined to be important beforehand. The self-reports seem to be in general agreement with the cue use data from the process descriptions. The policy-capturing models provide data about weighting that is not supplied by the process analysis. For instance, Fit was not the most frequently used, but when used, it may have been heavily weighted. Finally, the process-tracing analysis provides information about the task that enables us to better understand the poor showing of the policy-capturing models and to interpret the self-reports.

Discussion

This study has provided an opportunity to examine teacher judgment from multiple perspectives. Surprisingly, there are few examples of this kind of comparison in the judgment literature. (Einhorn, Kleinmutz, & Kleinmutz, 1979, and Yinger, 1975, are exceptions.) As a result, researchers know little about the comparative strengths of the various methods and the relative suitability of the methods for yielding certain kinds of information. Our intent in initiating the set of studies, of which this study is one part, was to compare policy-capturing and process-tracing methods "head-to-head" to determine

which was better. We have found, like others who have attempted this (see especially Einhorn et al., 1979) that this was a naive approach to the problem and that various methods provide unique and complementary results.

Method Characteristics

Policy capturing. These methods have been shown to provide high predictive power in a wide variety of judgment settings (see Slovic & Lichtenstein, 1971, and Slovic, Fischhoff, & Lichtenstein, 1977). It has also yielded high explanatory results as measured by goodness-of-fit criteria such as R^2 values. There is some debate, however, about whether the success of modeling human judgment with linear models is due to the linearity of judges' behavior or due to characteristics of linear models.

Dawes and Corrigan (1974), for example, made the observation that linear models have typically been applied in situations where the predictor variables are monotonically related to the criterion and where there is error in the independent and dependent variables. They showed that these conditions insure good fits by linear models, regardless of whether or not the weights in such models are optimal.

Self-reports. The degree to which a judge is (or can be) aware of the weight attached to his or her judgments has been the subject of considerable debate (see, for a review of this controversy, Slovic & Lichtenstein, 1971, and Slovic et al. 1977). More recently, the use of verbal reports as data has come under considerable criticism (Nisbett & Wilson, 1977). These criticisms have been convincingly rebutted by others (Ericsson & Simon, 1980).

Ericcson and Simon (1980) have developed an information-processing theory that specifies conditions under which verbal reports will be most reliable. Simply put, they argue that verbal reports will be most valid and reliable

when a subject is asked to report on the contents of short term memory--that which he or she is currently attending to. Inconsistent introspective reports will be more likely in response to probes that are too general to elicit the information actually sought or as a result of requests that require subjects to use inferential processes to fill out and generalize incomplete or missing memories (p. 247).

The type of self-report requested in this study is not of the level of reliability of in-process reports like "thinking aloud." It is a retrospective and somewhat generalized verbal report of the type that Ericsson and Simon would call a "general report." The fact that the self-reports were collected immediately after the task and requested information that was certainly a part of the judges' deliberations provide some support for the credibility of these verbal reports. At the most, the self-reports give some insight into the actual weightings incorporated by the teachers. At the least, they provide indications of the teachers' implicit theories about the various ways certain kinds of information are considered in making instructional decisions.

Process tracing. Process tracing has produced detailed models of problem solving and judgment in a variety of situations, both in the laboratory and in natural settings (see Shulman & Elstein, 1975). Recently research has demonstrated that process analysis can provide information about judgment that is unavailable through mathematical models. Einhorn et al. (1979) provide an anecdote demonstrating this point from a study of judgments of the nutritional quality of breakfast cereals.

> In the protocol, the judge uses the cue "calories" many time, yet, the cue receives no significant weight in the regression equation even though it is not very highly correlated with other cues. When one examines the variance of the cue in the sample of cereals used, the discrepancy becomes clear; namely, calories has a small variance and so it cannot receive a high weight in the regression equation. In contrast, the protocol indicates that the subject was paying attention to this cue. Whether the subject was really using this cue

> is problematic, since our results indicate that it could not have been of much help in discriminating between the brands. (In the extreme case, a cue with no variance should be irrelevant.) In any case, the process model clearly shows that one can attend to cues, and thus feel that one has used them, without such cues receiving significant weights in the regression analysis (p. 482).

Further support for the role of process tracing in interpreting the results of other data has been offered by Einhorn and Hogarth (1981).

> The use of weights in models as reflecting differential cue importance ignores the importance of attention in subjective weight estimates. . . Correspondence between subjective and statistical weights requires that people attend to and evaluate cues and that cues contain both variance and low intercorrelations. Disagreement between subjective and statistical weights can thus occur for three reasons: 1) people indeed lack insight; 2) people attend to, but cannot use, cues that lack variance (Einhorn et al., 1979), 3) cues to which attention is not paid are correlated with others such that the nonattended cues receive inappropriate statistical weights. Both process-tracing methods and statistical modeling are necessary to untangle these competing interpretations. (pp. 62-63)

Multi-method approaches. We join others in advocating the importance of approaching the study of human judgment using a variety of methods. No one method provides all the necessary information. As Einhorn et al. (1979) put it, "some process modelers may not be seeing the forest for the trees while some statistical modelers may not see any trees in the forest" (p. 483). While various approaches treat the underlying process at different levels of detail, each method can provide important data. Process analysis provides information about the judgment process and cue use. Policy-capturing models can provide information about the relative emphasis put on various cues. Self reports may provide confirmation of the information provided by the other methods, and depending upon the way in which probes are directed, reflect actual processes or the judge's beliefs about how these decisions should be made.

Factors Influencing Judgment

The task. Cognitive approaches to understanding human problem solving and decision making have characterized thinking as being adaptive to the situation at hand. Similiarly, researchers have found that the behavior of the person solving a problem or making a decision tells us as much or more about the structure of the task as about the unique characteristics of the person involved (Shulman & Elstein, 1975, p. 14). As a result, researchers who study decision making have increasingly turned their attention to trying to better understand and model task effects (see Einhorn & Hogarth, 1981).

Two factors related to task effects have become especially salient in the research presented in this paper. First, the nature of the task seems to strongly interact with the assumptions and suitability of the methods used to model judgment behavior. For example, policy-capturing models have proven to be most effective in situations where the objects to be judged are self-contained in terms of the information needed by the judge and where the judgment task requires little or no manipulation of the files prior to judgment. Examples of these as-given tasks include judging an admission file, and MMPI[3] profile, or a simulated student profile. These kinds of judgment tasks do not require that the judge use or plan to use the items in any real way. In contrast, judgment tasks such as those presented in this study imply a to-be-used criterion.

The activity descriptions used in this study were, in effect, plans for action--descriptions of what was to be done by a class. The might-be-done aspect seemed to require transformation of the activities, primarily by placing them in the context of the individual teacher and his or her students. As

---

[3]Minnesota Multiphasic Personality Inventory

a result of this mental manipulation, much additional information is brought to the activity descriptions and the teachers seemed to be judging not the activities-as-given but the activities-as-imagined-in-use (see Yinger & Clark, 1982 for more detail).

Therefore, the failure of the policy-capturing models may be partly a function of the task framework. An examination of the judgment literature indicates that mathematical models seem to be most successful in judgment situations that are self-contained--situations where objects may be judged solely as given. Studies that have been less successful in representing judgment with these models have typically used more complex stimuli (e.g., written descriptions vs. numerical profiles) and have implied an in-use criterion (see for example, Borko, 1978; Cone, 1978; Russo, 1978; and Floden et al., 1981).

The second consideration related to task is the apparent trade-off between control and representativeness of the judgment task. As mentioned above, mathematical models seem to be best suited to laboratory tasks that can be simplified and controlled. As tasks more closely resemble real-life judgments, they become more difficult to model using these methods. Einhorn and Hogarth (1981) in discussing this problem refer to the work of Ebbesen and Konecni (1980) who have studied several judgment tasks in both laboratory and natural settings (for example, setting of bail and driving a car) and have found major differences in results. Einhorn and Hogarth cite Ebbesen and Konecni's conclusions:

> There is considerable evidence to suggest that the external validity of decision making research that relies on laboratory simulations of real-world decision problems is low. Seemingly insignificant features of the decision task and measures cause people to alter their decision strategies. The context in which the decision problem is presented, the salience of alternatives, the number of cues, the concreteness of the information, the order of presentation, the similarity of cue to alternative, the nature of the decomposition, the form of the measures, and so on, seem to affect the decisions that subjects make. (Einhorn & Hogarth, 1981, p. 81)

Experience. Among cognitive researchers, the nature and the effects of experience have become a popular topic of research and theory. Much of this work has been devoted to the differences between novices and experts in various task environments and the factors contributing to an eventual shift from novice to expert.

There are two findings from this research that suggest that the experience level of a judge should be taken into consideration in interpreting the results of various modeling methods. First, research indicates that experts are more likely than novices to recognize (perceive/understand) and represent problems using large-scale functional units (e.g., schemas, scripts, routines) that focus on the crucial underlying structure and components of the problem (de Groot, 1965; Hinsley, Hayes, & Simon, 1977; Larkin, 1979; Newell & Simon, 1972). The reliance on these large units of knowledge and skill suggest that judgment tasks may activate large pieces of knowledge and experience that immediately become part of the judgment task. Researchers have also found that experts are more likely than novices to mentally simulate action prior to its execution by means of incorporating complex and detailed representations of action within a particular environment (de Groot, 1965; Jeffries, 1982; Larkin, 1979).

These findings suggest that the more experience a person has, the more likely he or she is to embellish and transform the information provided in the judgment task. Modeling methods such as policy capturing that assume knowledge and control of the content of the objects being judged may be less accurate for experienced judges. In comparison, we would hypothesize that novice judges, having less stored knowledge and experience to draw upon and incorporate into their judgments, would rely more exclusively on the information presented in the task. In this later case, mathematical models would be expected to have a better fit.

A study by Slovic, Fleissner, and Bowman (1972) provides results that are consistent with these predictions. In a study of investment decision making, they found a negative correlation (-.43) between years of experience and self-insight. Self-insight was calculated by correlating a broker's subjective weights (self-reported) with his calculated effects (mathematical model). In other words, the more experienced the broker, the less agreement there was between his or her self-reported policy and that generated by the policy capturing.

Slovic and his colleagues interpreted these results as possibly suggesting that "the most experienced analysts produce verbal rationales for their evaluations that are less trustworthy than those of their inexperienced colleagues" (p. 300). Based on our discussion above, we would interpret these results as suggesting the inability of the mathematical model to represent accurately what the judge is actually doing.

Conclusions

Based on this study and the research presented in the discussion, we offer four conclusions as hypotheses to be considered in future research on teacher judgment.

First, we think that judges have better insight into their own decision processes than researchers have typically given them credit for. Researchers need to pay closer attention to the differences in language and level of detail offered by various methods. Researchers also need to evaluate carefully what kind of data will be used as a criterion to evaluate the validity of verbal reports.

Second, the form and complexity of the judgment task must be considered in evaluating the results of various modeling methods. Researchers need to develop better models of the tasks in which judgment is being examined.

Third, the more experienced the judge, the more likely he or she is to be judging elaborated and transformed mental models rather than the objects-as-given. Researchers need to know more about how experience influences judgment.

Fourth, one method of modeling judgment is not better than the others for all purposes. The three sources of data used in this study each provided different, though complementary information. Accurate descriptions of teacher judgment will be more likely if multi-method approaches are employed by researchers.

References

Anderson, B. (1977). Differences in teachers' judgment policies for varying numbers of verbal and numerical cues. Organizational Behavior and Human Performance, 19, 68-88.

Borko, H. (1978, March). An examination of some factors contributing to teachers' preinstructional classroom organization and management decisions. Paper presented at the annual meeting of the American Educational Research Association, Toronto.

Borko, H. & Niles, J. (1982, March). Factors contributing to teachers' decisions in the planning of reading instruction. Paper presented at the annual meeting of the American Educational Research Association, New York.

Clark, C.M., & Yinger, R.J. (1977). Research on teacher thinking. Curriculum Inquiry, 7, 279-304.

Clark, C.M., Yinger, R.J., & Wildfong, S.C. (1978). Identifying cues for use in studies of teacher judgment (Research Series No. 23). East Lansing, MI: Institute for Research on Teaching, Michigan State University.

Cone, R. (1978, March). Teachers' decisions in managing student behavior: A laboratory simulation of interactive decision-making by teachers. Paper presented at the annual meeting of the American Educational Research Association, Toronto.

Dawes, R.M. & Corrigan, B. (1974). Linear models in decision making. Psychological Bulletin, 81, 96-106.

de Groot, A.D. (1965). Thought and choice in chess. The Hague: Mouton.

Ebbesen, E.B., & Konecni, V.J. (1980). On the external validity of decision-making research: What do we know about decisions in the real world? In T.S. Wallsten (Ed.), Cognitive processes in choice and decision behavior. Hillsdale, NJ: Lawrence Erlbaum.

Einhorn, H.J., & Hogarth, R.M. (1981). Behavioral decision theory: Processes of judgment and choice. Annual Review of Psychology, 32, 53-88.

Einhorn, H.J., Kleinmutz, D.N., & Kleinmutz, B. (1979). Linear regression and process-tracing models of judgment. Psychological Review, 86, 465-485.

Ericcson, K.A., & Simon, H.A. (1980). Verbal reports as data. Psychological Review, 87, 215-251.

Floden, R.E., Porter, A.C., Schmidt, W.H., Freeman, D.J., & Schwille, J.R. (1981). Responses to curriculum pressures: A policy-capturing study of teachers' decisions about content. Journal of Educational Psychology, 73, 129-141.

Forte, I., Frank, I., & MacKenzie, J. (1973). Kids stuff: Reading and language experiences. Intermediate-junior high. Nashville, TN: Incentive Publications.

Hinsley, D.A., Hayes, J.R., & Simon, H.A. (1977). From words to equations: Meaning and representation in algebra word problems. In P.A. Carpenter and M.A. Just (Eds.), Cognitive processes in comprehension. Hillsdale, NJ: Lawrence Erlbaum.

Jeffries, R. (1982). A comparison of debugging behavior of expert and novice programmers. Paper presented at the annual meeting of the American Educational Research Association, New York.

Larkin, J.H. (1979). Information processing models and science instruction. In Lockhead, J. and Clement, J. (Eds.), Cognitive Process Instruction. Philadelphia, PA: Franklin Institute Press.

Newell, A., & Simon, H.A. (1972). Human problem solving. Englewood Cliffs, NJ: Prentice Hall.

Nisbett, R.E., & Wilson, T.D. (1977). Telling more than we can know: Verbal reports on mental processes. Psychological Review, 84, 231-259.

Russo, N.A. (1978, March). Capturing teachers' decision policies: An investigation of strategies for teaching reading and mathematics. Paper presented at the annual meeting of the American Educational Research Association, Toronto.

Shavelson, R.J., & Stern, P. (1982). Research on teachers' pedagogical thoughts, judgments, decisions, and behavior. Review of Educational Research, 51, 455-498.

Shulman, L.S., & Elstein, A.S. (1975). Studies of problem solving, judgment, and decision making. In F.N. Kerlinger (Ed.), Review of research in education (Vol. 3), Itaska, IL: F.E. Peacock.

Slovic, P., Fleissner, D., & Bauman, W.S. (1972). Analyzing the use of information in investment decision making: A methodological proposal. Journal of Business, 45(2), 238-301.

Slovic, P., Fischhoff, B., & Lichtenstein, S. (1977). Behavioral decision theory. Annual Review of Psychology, 28, 1-39.

Slovic, P., & Lichtenstein, B. (1971). Comparison of Bayesian and regression approaches to the study of information processing in judgment. Organizational Behavior and Human Performance, 6, 649-744.

Yinger, R.J. (1975). The diagnosis of sore throat children: A comparison of two judgment models. Unpublished paper, Michigan State University.

Yinger, R.J. (1977). A study of teacher planning: Description and theory development using ethnographic and information processing methods. Unpublished doctoral dissertation, Michigan State University.

Yinger, R.J., & Clark, C.M. (1982). Understanding teachers' judgments about instruction: The task, the method, and the meaning (Research Series No. 121). East Lansing, MI: Institute for Research on Teaching, Michigan State University.

Yinger, R.J., Clark, C.M., & Mondol, M.M. (1981). Selecting instructional activities: A policy capturing analysis (Research Series No. 103). East Lansing, MI: Institute for Research on Teaching, Michigan State University.